MODAL LOGIC AND AGENCY PART 1 - MODAL SYNTAX, SEMANTICS, AND PROOF THEORY

Alessio Lomuscio

Department of Computer Science
King's College London

alessio@dcs.kcl.ac.uk

Thanks to Ian Hodkinson for some of the material in earlier drafts of these notes.

Overview

- Why modal logic for MAS.
- Syntax of Modal Logic
- Semantics.
- Model Construction.
- Satisfaction, Validity.
- Basic sound and completeness concepts.

This course - what it is

The field of the applications of modal logic to agency is considerable, and it would not be reasonable to be aiming at giving more than an overview of the basic techniques and applications. What we shall do is:

- Study basic syntactical, semantical, and proof-theoretical notions of modal logic.
- Survey basic concepts of epistemic logic (logic of knowledge), and more briefly doxastic logic (logic of belief).
- Present a few systems in which several modalities interact with one another to specify several aspects of agency.
- See one alternative semantics - interpreted systems -

This course - what it is not

- An in-depth course on modal logic.
- A course on verification of MAS. We take inspiration from these formalisms, but how we may go about verifying a system specified in these frameworks is not discussed here.

Unfortunately we do not have time to discuss these issues and we refer to specialised literature.

Textbooks for further reading

- R. Goldblatt. *Logics of Time and Computation*. Second edition. CSLI Publications, 1992. Terse and authoritative. Covers most of the “non-epistemic parts” of this course.
- Fagin, Halpern, Moses, Vardi. *Reasoning about knowledge*, MIT Press 1995. The Reference for anyone doing research on interpreted systems semantics and their application.
- van der Hoek, JJ Meyer. *Epistemic Logic for AI and Computer Science*, CUP. Comprehensive and gentle on Epistemic Logic.
- M. Wooldridge. *Reasoning about Agents*. Modal formalisms for agents.

Prerequisites for this course
We assume working knowledge of Propositional calculus, and its basic constructs.

Modal logic - Introduction

Modal logic offers:

- A natural way of handling the concepts of necessity, possibility, temporal flows, knowledge, belief, etc.
- A clear and natural semantics.
- A heritage of techniques for proving meta-logical results about modal systems, such as decidability, completeness, computational complexity, etc.

All these characteristics make it one of the formalisms of choice when modelling particular computing systems.
Modal logic is particularly used in the field of multi-agent systems (open distributed systems in which entities are autonomous in nature). We shall see some applications of these.

Syntax, semantics

It is always useful to separate two concepts: syntax and semantics.

- *Syntax* has to do with how we write formulas. It defines the logical language that we use when writing formulas.
- *Semantics* focuses on giving an *interpretation* to formulas constructed according to the syntax.

Definition 1 (Syntax) *Assume a (countably infinite) set of propositional variables P, the set $\mathcal{L}$ of valid formulas of propositional modal logic is inductively defined by:*

- *any* $p \in \mathcal{L}$.
- **true** $\in \mathcal{L}$.
- *If* $\varphi, \psi \in \mathcal{L}$*, then* $\neg\varphi, \varphi \wedge \psi, \Box\varphi \in \mathcal{L}$.
- *Nothing else is in* $\mathcal{L}$.

We write $\varphi, \psi, \ldots$ for arbitrary formulas in $\mathcal{L}$, and $p, q, \ldots$ for atoms in P. **true** is the symbol for logical truth, the unary connector $\neg$ stands for logical not, while the binary connector $\wedge$ stands for logical and.
According to Definition 1, the set $\mathcal{L}$ of the formulas we shall use is an extension of the propositional calculus by an operator, $\Box$, that we shall call *modality*, or *box*.

Abbreviations

It is sometimes convenient to use other connectives that can be defined from Definition 1.

- **false** stands for $\neg$**true**.
- $\varphi \vee \psi$ stands for $\neg(\neg\varphi \wedge \neg\psi)$.
- $\varphi \rightarrow \psi$ stands for $\neg\varphi \vee \psi$, i.e., $\neg(\varphi \wedge \neg\psi)$.
- $\Diamond\varphi$ stands for $\neg\Box\neg\varphi$.

Most of these abbreviations are known from propositional logic. $\wedge$ stands for logical and, while $\rightarrow$ stands for logical implication. The modal operator $\Diamond$ is the dual of $\Box$ and is read as "diamond".

Modalities - readings

Although we can develop the whole of modal theory without any reference to the intuitive meaning of the modalities, this is not the way it was developed. Indeed as it will become clearer throughout the course, different flavours of modal logic were developed depending on the intuitive application. Here we shall focus on some meanings of the modal box, others exist.

Meaning of $\Box$	**Reading of $\Box\varphi$**	**Reading of $\Diamond\varphi$**
Necessitation	It is necessary that φ	It is possible that φ
Epistemic	It is known that φ	It is considered (epistemically) possible that φ
Doxastic	It is known that φ	It is considered (doxastically) possible that φ
Deontic	It is obligatory that φ	It is possible that φ
Provability	It is provable that φ	It is consistent that φ
Temporal	For ever in the future φ	Sometime in the future φ

Other examples:

Epistemic	$\Box\Box\varphi$	φ is known to be known
Epistemic	$\Box\Diamond\varphi$	φ is known to be regarded as possible
Temporal	$\varphi \rightarrow \Box\varphi$	If φ is true now, it will always be true

Note the expressive power of combination of operators with different interpretations! For example $\Box_T \Box_K \varphi$ reads as "the agent will forever know that φ" if we give a temporal reading to the first modality and an epistemic one to the second. We shall explore some of these aspects towards the end of the course.

Possible worlds semantics

Given Definition 1 we now know how to construct formulas of the logical language. Let us now discuss the concept of truth. Given a situation, or a system, some formulas will be true, others will be false. For example "it will rain forever" will be false in a situation in which one day of sun follows in a week. Other formulas will happen to be true, perhaps "it is sunny tomorrow". These will be called "contingent" truths. Other formulas seem *necessarily true* (or *necessarily false*). For example, consider a temporal interpretation of the modalities, and the formula: $\Box p \rightarrow \Box\Box p$, where p reads as "it rains". The formula reads "if it rains forever, then it rains forever forever", which seems in line with our intuition of temporal flows of time.
So we need to be careful with different notions of truth. Our approach will be to express these concepts in a formal way. This is what we do in the following. We abstract from different readings of the modalities and present formal definitions.

Kripke Frames

A good semantics will be easy and intuitive, but at the same time it will capable of accommodating all the different readings of the modalities. There is not *one* semantics for modal logic, and indeed the matter has received a lot of attention in the past. What we present here is the "usual" possible worlds semantics as defined by Kripke in the late 50s (based on ideas from Tarski and Johnson of early 50s). Further reading on the matter can be found in Kripke, Naming and necessity, Harvard Press, 1980.
The basic idea of *possible worlds semantics* (or *Kripke semantics*) is that a formula which is "possibly true" is true at some worlds, and perhaps false at others. Let us be formal though.

Definition 2 (Frame) *A* Kripke frame F *is a pair* $F = (W, R)$*, where* $W \neq \emptyset$ *is a set of possible worlds, and* $R \subseteq W \times W$ *a relation defined on these.*

We shall indicate elements of W as w, etc. These are also called states, or points.

Exercise 1 *Define Kripke frames of some mathematical (and not) structures you are familiar with.*

For example:

- (N, suc), the set of natural numbers with the relation of successor is a Kripke frame.
- $(\{w\}, \emptyset)$ is a frame.

- (Z, suc), the set of relative numbers with the relation of successor is a Kripke frame.
- $(N, <)$, the set of natural numbers with the relation of less than is a Kripke frame.
- Any finite and infinite tree is a frame.

Kripke frames are unvalued structures, i.e., we cannot evaluate formulas on them. To do this, we need Kripke models.

Kripke Models

Definition 3 (Model) *A* Kripke model M *is a pair* $M = (F, \pi)$, *where* $\pi : P \to 2^W$ *is a valuation for the atoms*[1].

Intuitively $\pi(p) = \{w_1, w_2\}$ represents the fact that p is *true* at w_1, w_2 and *false* at $W \setminus \{w_1, w_2\}$.
By slight abuse of notation we shall often indicate a model M as a triple $M = (W, R, \pi)$, where (W, R) is its *underlying frame*.
We now have all ingredients in place to define formally satisfaction of a formula of the logical language of Definition 1.
Satisfaction on Kripke structures is defined as follows:

Definition 4 (Satisfaction) *The satisfaction of a formula* $\varphi \in \mathcal{L}$ *at a world* $w \in W$ *of a model* M, *formally* $M \models_w \varphi$, *is inductively defined as follows:*

$$\begin{array}{lll} M \models_w \mathbf{true} & & \\ M \models_w p & \textit{if} & w \in \pi(p), \\ M \models_w \neg\varphi & \textit{if} & M \not\models_w \varphi, \\ M \models_w \varphi \wedge \psi & \textit{if} & M \models_w \varphi \textit{ and } M \models_w \psi, \\ M \models_w \Box_i \psi & \textit{if} & \textit{for each } w' \in W \textit{ we have } wR_iw' \\ & & \textit{implies } M \models_{w'} \psi. \end{array}$$

We read $M \models_w \varphi$ *as "*φ *is true at* w *in model* M*".*

Satisfaction for the other logical connectives can be defined from the abbreviations.

[1] Given a set A, 2^A is the power set of A, i.e., the set containing all subsets of A. For example if $A = \{1, 2\}$, then $2^A = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.

Exercise 2 *Check that the abbreviations introduced above give the intended results. For example that:*
$M \models_w \varphi \vee \psi$ *if* $M \models_w \varphi$ *or* $M \models_w \psi$

Exercise 3 *Consider the following Kripke model, built on a set of propositional variables* $P = \{p, q\}$. $M =$
$(\{w_1, w_2, w_3, w_4\}, \{(w_2, w_1), (w_3, w_2), (w_4, w_2), (w_4, w_3), (w_3, w_3)\}, \pi)$, *where* π *is defined by* $\pi(p) = \{w_2, w_4\}; \pi(q) = \{w_2, w_3, w_4\}$.
Work out whether or not the following hold.
$M_{w_4} \models \Box p$, $M_{w_4} \models \Diamond p$, $M_{w_3} \models \Box p$, $M_{w_3} \models (\Diamond q \to \Diamond\neg p) \to \Diamond(\neg p \wedge \neg q)$, $M_{w_2} \models \Box\bot$, $M_{w_2} \models \Diamond\bot$, $M_{w_1} \models \Box\mathbf{true} \wedge \Box\mathbf{false}$, $M_{w_2} \models \Diamond\mathbf{true} \vee \Diamond\mathbf{false}$.

Check whether or not the following formulas hold.

Validity

Satisfaction of formulas at a point of a model corresponds to our notion of contingent truth. Still, there are formulas that are true at all points of a model: these in a way *characterise* the model.

We can go further. Given a frame F, and any interpretation π on it, there are formulas that are true *irrespective* of the interpretation function we give(!). This is a remarkable property, because it means that such formulas represent structural properties of the frame, that can be expressed syntactically via a modal language. We shall soon find out more about what these formulas are and how they are related to properties on the frames, but for now let us be content to say that they exist. For example the formula $\Box$**true** is "true at any frame", because if we consider any interpretation built on any frame, it will be satisfied at any point. Let us make this all more precise.

Validity on Kripke structures is defined as follows:

Definition 5 (Validity) *A formula φ is valid on a model $M = (W, R, \pi)$, formally $M \models \varphi$, if for any point $w \in W$ we have $M \models_w \varphi$.*

A formula φ is valid on a frame $F = (W, R)$ if for any interpretation π we have $(F, \pi) \models \varphi$.

A class of models $\mathcal{M}$ validates a formula φ if for any model $M \in \mathcal{M}$ we have $M \models \varphi$.

A formula φ is valid on a class of frames $\mathcal{F}$ if for any frame $F \in \mathcal{F}$ we have $F \models \varphi$.

Validity - examples

Exercise 4
- *Consider the frame $F = (\{w\}, \{(w, w)\})$. Are there any formulas that are valid on this frame?*
- *Consider the class of all frames. Can you think of any formulas that are valid on any frame of the class, i.e., that are valid on the class?*
- *Consider the class of reflexive frames, i.e. frames whose relation is reflexive (for all w, wRw). Are there any formulas that are valid on this class?*

A hyrerarchy of truth

Note that there is a hyrarchical chain of quantification that connects semantical notions from the one of truth at a point of a model to the one of validity on a class of frames.

$\mathcal{F} \models \varphi$	Validity on classes of frames
$F \models \varphi$	Validity on a frame
$(F, \pi) \models \varphi$	Validity on a model
$(F, \pi) \models_w \varphi$	Truth on a point of a model

From the first to the second definition the difference is $\forall F \in \mathcal{F}$. The second differs from the third by $\forall \pi$, while between the third and the fourth there is $\forall w \in W$.

Hilbert systems - introduction

So far we have explored the importance of semantics in modal logic. In particular we have seen how to give an interpretation to modal formulae,
All this has to do with the *s*emantics of modal logic.
But there is another side, one that traditionally was developed first in fact. It has to do with syntax, or proof-theory. The basic concepts there are the ones of *Logic, and Axiomatisation*.

Logics

We know how to define the syntax of modal logic — this is essentially an extension of propositional calculus by means of one (or more) modalities. BNF syntax defines the *whole language* under consideration, but suppose we intend to spell out the properties that a particular (computing) system enjoys.
Presumably to do this we need to use only a subset of the whole set of valid formulas. The concept of logic is useful in this case:

Definition 6 *Given a (modal) language $\mathcal{L}$, a logic L is a subset $L \subseteq \mathcal{L}$. Elements $l \in L$ are called* theorems *of L.*

Note that the definition of logic goes beyond computing and is used in many other different fields. For example consider the language of first-order logic as it is used to represent the set of meaningful (in the sense of well-formed, not in the sense of valid!) formulas about arithmetics. One logic of this set is the set of the valid formulas in arithmetics.
What happened to the concept of *truth*?
Truth is a semantical concept - here we have the concept of consistency.

Definition 7 *A logic L is consistent if $\perp \notin L$. A logic L is inconsistent if it is not consistent.*

Examples of logics:
$\{p, q\}, \{\Box p, \textbf{false}, \textbf{true}\}, \{\textbf{true}, \Box p, \Box\Box p, \Box\Box\Box p, \ldots\}, \ldots$
Notation: we write $L \vdash \varphi$ to mean that φ is a theorem of L, i.e., $\varphi \in L$.

Axiomatisation

The concept of logic is useful but a little cumbersome to work with. Consider the countably infinite set:

$$L = \{\Box p, \Box\Box p, \Box\Box\Box p, \Box\Box\Box\Box p, \ldots\}$$

Intuitively we could alternatively represent it by saying that

- $\Box p \in L$,
- If $\varphi \in L$, then $\Box\varphi \in L$.
- Nothing else is in L.

Exercise 5 *Prove that the two formulations are equivalent.*

This is the essence of the concept of *axiomatisation*: intuitively we would like to have a set of starting formulae and some rules that would generate other formulas. The former go under the name of *axioms*, the latter under the name of *inference rules*.

Definition 8 (Axiomatisation) *Given a language $\mathcal{L}$, a finite axiomatisation $AX = (A, R)$ is a pair where $A = \{a_1, \ldots, a_n\}$, $a_i \in \mathcal{L}$, $i = 1, \ldots, n$ is a set of axioms, and $R = \{r_1, \ldots r_m\}$ is a set of rules of the form: $\vdash \varphi \Rightarrow \vdash \psi$, where $\varphi, \psi \in \mathcal{L}$.*
A theorem of AX is either an element of A, or a formula deducible from A by means of rules in R, or a formula deducible from another theorem by means of rules in R.

Definition 9 *Given a language $\mathcal{L}$, a logic $L \subseteq \mathcal{L}$ is finitely axiomatisable if there exists an axiomatisation $AX = (A, R)$ such that*

- *For any $\varphi \in L$, L is a theorem of AX,*
- *For any theorem φ of AX, $\varphi \in L$.*

Note1: Intuitively we would like to use the concept of axiomatisation in place of the concept of logic because of its convenience. Indeed, in the following we shall assume that all logics we deal with are finitely axiomatisable. Indeed for the purposes of this course we shall use the terms “logic” and “axiomatisation” as synonyms. But it is important to realise that in general these are different concepts.
Note2: We could extend the definition above to deal with infinite axiomatisations; indeed there are logics that are not finitely axiomatisable but are axiomatisable. We do not deal with these in this course.

Systems of Modal Logic

Given the compact way we have found to express logics, we can now study and compare different systems of modal logic, i.e., different logics, or axiomatisations.
Let us first analyse the following system, called system K:

Definition 10 (System K)
- *All valid Propositional Calculus formulae* **PC**,
- $\Box(p \to q) \to (\Box p \to \Box q)\mathbf{K}$,
- $\vdash \varphi \Rightarrow \vdash \varphi[\psi_1/p_1, \ldots, \psi_n/p_n]\mathbf{US}$
- $\vdash \varphi, \vdash \varphi \to \psi \Rightarrow \vdash \psi\mathbf{MP}$,
- $\vdash \varphi \Rightarrow \vdash \Box\varphi\mathbf{NEC}$

System K is a very important system as we shall discuss later on. For the moment let us analyse the system in more detail.
Axiom PC guarantees that all propositional tautologies are included in the system. All systems that include axiom PC are called “classical”.
The peculiarity of Axiom K will become clear later on.
The inference rule US guarantees that if a formula φ is provable in the system, then so is the formula resulting from replacing uniformly any variable p_i occurring in φ with any well-formed formula ψ. For example if $\vdash \Box(p \to q)$, then $\vdash \Box(\Box(r \vee s) \to q)$, where we replace p with $\Box(r \vee s)$ (note that in this case p only occurs once in φ, if it occurred more than once, we would be forced to substitute all instances of it).
The inference rule MP, or Modus Ponens, is the classical inference rule of propositional calculus.
The inference rule NEC guarantees that we can obtain another theorem by boxing a theorem.

Example 1 *Interpret system K in view of a temporal, epistemic, and doxastic interpretation of the modalities and evaluate whether or not K is an appropriate model for these concepts.*

Normal Modal Systems

The importance of system K is more theoretical than in terms of applications. As it happens system K is seldom used as a model for key properties of temporal evolution or epistemic properties. What is important of system K is that it provides an axiomatisation for the smallest “normal” logic.

Definition 11 (Normal systems) *A logic L is normal if it includes system K.*

We shall soon discover why system K is so important. To anticipate, it represents the smallest set of formulae that are valid on the whole class of Kripke frames.
Before we do this, let us analyse other normal modal systems. Given that by Definition 11 all normal systems include K, we may define other normal systems by adding either axioms or inference rules to system K (or both). Although other possibilities exist, it is customary to consider other normal systems as:

$$L = K \cup \varphi,$$

where L is the normal system to be specified, K is system K, and $\varphi \in \mathcal{L}$ is formula (possibly a conjunction of formulas.

Definition 12

- $T = K \cup \{\Box p \rightarrow p\}$
- $D = K \cup \{\Box p \rightarrow \Diamond p\}$
- $4 = K \cup \{\Box p \rightarrow \Box\Box p\}$

Other systems can be defined by extending the systems above. For example

Definition 13

- $S4 = T \cup 4$
- $S5 = T \cup \{\Diamond p \rightarrow \Box\Diamond p\}$
- $4 = K \cup \{\Box p \rightarrow \Box\Box p\}$

Completeness - Introduction

So far we have seen many different variations on the concept of “truth”:

- truth of a formula at a world of a model,
- truth of a formula on a model,
- truth of a formula on a frame,
- truth of a formula on a class of frames,

Each of these is more general, and culminates with the concept of *validity on a class of frames*: $\mathcal{F} \models \varphi$.
We have also defined Hilbert systems. There, truth means *provability* from a set of axioms via some inference rules: $S \vdash \varphi$.
Question: Is there any relation between validity on a class of frames, and provability on a system? More precisely, can we find particular classes of frames and particular Hilbert systems in which these two relations are strongly related? This is the questions that soundness and completeness address.

$$\mathcal{F} \models \varphi \Leftrightarrow S \vdash \varphi$$

dcs.kcl.ac.uk — 2002 Modal Logic and Agency

Why is it important??

Many reasons:

- To compare different logic systems: before completeness results became known deciding whether or not two systems S_1, S_2 were equivalent was a very nontrivial matter. Thanks to completeness we can reason about the corresponding semantical structures.
- To validate computing systems. If we have a specification given semantically, by using completeness results we can reason about it via proof-theoretical tools, automatic theorem provers for example. Conversely, if we are given a syntactical specification of a computing system, completeness guarantees that we can reason about its properties by using modal theoretical tools, e.g., model checkers.

Soundness and completeness of a logic with respect to a class of frames opens the possibility of choosing the most appropriate tool for the task. We will see an example of this when we compare different axiomatic systems, and when we study how to represent knowledge bases by a modal formalism.

© A. Lomuscio Lecture 1 - The basics

Soundness

Given $\mathcal{F}$, and S, solving one direction of the problem is relatively easy: $S \vdash \varphi \Rightarrow \mathcal{F} \models \varphi$.
If system S is *normal* it can be expressed as $S = K \cup \{\Lambda\}$, i.e. axiom K, the inference rules of MP, and US, and a set of specific axioms Λ. Since we know that K is sound with respect to any class of frames, in order to prove soundness, all we need to prove is that Λ is valid on the class $\mathcal{F}$.

Lemma 1 *System $T = K \cup \{\Box p \to p\}$ is sound with respect to the class of reflexive frames.*

Proof: All we need to show is that axiom T is valid on reflexive frames. So consider a model $M = (F, \pi)$ built on any reflexive frame F by a valuation π. Suppose $M \models_w \Box p$. So for any wRw' we have that $M \models_{w'} p$. But R is reflexive, so wRw, and so $M \models_w p$. Therefore we have $M \models_w \Box p \to p$. ▫

Completeness of basic systems

There are many techniques to prove completeness. The easiest to use is what is called the *canonical model technique*. We do not present any of these here, but just give some of the results that can be obtained.

Theorem 1 *For the table below the following holds: $L \vdash \varphi$ if and only if π is valid on the corresponding class of Kripke frames.*

Logic	**Structural property**
K	*Arbitrary frames.*
$K \cup \{\Box p \to p\}$	*Reflexive frames.*
$K \cup \{p \to \Box\Diamond p\}$	*Symmetric frames.*
$K \cup \{\Box p \to \Box\Box p\}$	*Transitive frames.*
$K \cup \{\Box p \to \Diamond p\}$	*Serial frames.*
$S5$	*Equivalence frames*

Conclusions

We have seen the basics concepts of modal logic:

- Syntax,
- Kripke Semantics,
- Axiomatisations,
- Soundness and completeness.

In the following we shall try to apply these to model aspects of agency, such as knowledge, beliefs, etc.